

Linguistically-based informational masking in preschool children

Rochelle S. Newman, Giovanna Morini,^{a)} and Faraz Ahsan

Department of Hearing and Speech Sciences, University of Maryland, 0100 Lefrak Hall,
College Park, Maryland 20742, USA

rnewman1@umd.edu, gmorini@umd.edu, farazahsan95@gmail.com

Gerald Kidd, Jr.

Department of Speech, Language and Hearing Sciences, Boston University,
635 Commonwealth Avenue, Boston, Massachusetts 02215, USA

gkidd@bu.edu

Abstract: Previous work has shown that young children exhibit more difficulty understanding speech in the presence of speech-like distractors than do adults, and are more susceptible to at least some form of informational masking (IM). Yet little is known about how/when the “susceptibility” to linguistically-based IM develops. The authors tested adults, school-age children (aged 8 yrs), and preschool-age children (aged 4 yrs) on sentence recognition in the presence of normal speech, “jumbled” speech, and reversed speech distractors. As has been found previously with adults [e.g., Summers and Molis (2004). *J. Speech, Lang. Hear. Res.* **47**, 245–256], children in both age groups showed a release of masking when the distractor was uninterpretable (reversed speech). This suggests that children already demonstrate linguistically-based IM by the age of 4 yrs.

© 2015 Acoustical Society of America

[Q-JF]

Date Received: January 25, 2015 Date Accepted: May 12, 2015

1. Introduction

The ability to understand speech in the presence of distracting sounds (speech or otherwise) is a critical skill for all listeners, but perhaps is especially important for preschool and school-age children, who spend significant portions of time in classroom settings where noise levels can be high. Distractors can interfere with their ability both to comprehend language and learn from it (e.g., Hétu *et al.*, 1990; Hygge *et al.*, 2002; Shield and Dockrell, 2003). Children have more difficulty understanding speech in the presence of background sounds than do adults, particularly when the distractor consists of speech (Fallon *et al.*, 2000; Hall *et al.*, 2012; Klatte *et al.*, 2010). However, the degree of interference caused by unwanted speech may depend on its linguistic content. If so, the masking young children experience from unwanted speech may develop over time as they mature and acquire language, and may differ among different age groups. The present study explores this age-related difference.

Background sounds can have effects at multiple levels. First, unwanted sounds may obscure the target speech due to overlap in the neural representations in the auditory periphery, a factor known as “energetic masking” (EM). This can make portions of the target signal inaudible, degrading the quality of the speech. But background distractors can also cause “informational masking” (IM), originating at higher levels of processing, which is the focus of the present study. IM may be caused by multiple aspects of the target and competing sounds, or by the way those sounds are presented (see Kidd *et al.*, 2008). But when the competing and target sounds are both speech, linguistic factors may be key (Brouwer *et al.*, 2012).

Adults show greater masking from a background speech signal that is intelligible (such as forwards speech in a known language) than from an unintelligible speech or speech-like signal (e.g., reversed speech or foreign speech; see Francart *et al.*, 2011; Hoen *et al.*, 2007; Summers and Molis, 2004). Summers and Molis suggested this was the result of “competitive processing of words and phrases within the speech masker” (p. 245).

Because language develops from infancy throughout childhood it seems likely that this linguistically-based IM would also change and develop over time. Although

^{a)}Present address: School of Education, University of Delaware, Newark, DE 19716, USA.

preschool-age children have been found to be more vulnerable to IM caused by stimulus uncertainty in nonspeech sounds than adults (Oh *et al.*, 2001), little is known about children's susceptibility to linguistically-based IM. Children's language skills are still developing, and they may not be able to process the linguistic message of a masker in the same manner as an adult, and so it is possible that they might in fact show less linguistically-based IM than adults. In line with this idea, Newman (2009) has found that young infants show no apparent difference between a forwards- and backwards-speech masker, presumably because they are less likely to interpret either signal linguistically. Also, Johnstone and Litovsky (2006) reported that the release from masking due to time reversal of speech maskers observed for adults was not present for children aged 5–7 yrs who exhibited about the same performance for each masker type. While these findings suggest that some developmental change occurs that affects IM, there is little evidence to date that indicates when such a change might take place.

The current study was intended to explore this issue more directly. We examined the identification of target words in the presence of three maskers that had equal spectral overlap (long-term average spectra) with the target (i.e., same EM), but which differed in their potential for producing linguistically-based IM. We tested three groups of listeners: college-age young adults, 8-yr-old school children, and 4-yr-old preschool children. These two groups of children provide insight into the development of linguistically-based IM that must occur from infancy (e.g., Newman, 2009) through adulthood (e.g., Kidd *et al.*, 2008). Our hypothesis was that meaningful and non-meaningful speech distractors would produce similar masking in the younger children (that is, they would not show linguistically-based IM) if their language development was immature, whereas older children and adults would show significantly less masking for the non-meaningful maskers (an indication of additional masking caused by an intelligible linguistic message in the background). Based on the results of the limited work on this topic reported in the past, our prediction was that this developmental change would occur at some point between 4 and 8 yrs of age.

2. Methods

Twenty participants aged 4–4.5 yrs (“preschool children”; 14 female, 6 male, range 4 yrs 0 months–4 yrs 6 months), 20 participants aged 8.5–9 yrs (“school-age-children”; 10 female, 10 male; range 8 yrs 6 months–9 yrs 1 month.), and 16 adults (14 female, 2 male) participated. All were native speakers of English, although 7 participants (4 preschoolers, 1 school-age, and 2 adults) heard another language 20% of the time or less. Data were dropped from an additional 16 participants (11 preschool children, 2 school-age children, 3 adults) for the following reasons: experimenter error ($n=1$); unwillingness to perform the task (e.g., pointing to multiple answers, not attending; $n=5$); incorrect age at test ($n=2$); refusal to wear headphones ($n=1$); failure to complete ($n=3$); diagnosis of attention deficit disorder ($n=1$); or video recording difficulties ($n=3$). All parents reported normal hearing, language, and cognitive development for their children, as well as no current ear infections, and all adults reported likewise.

Stimuli consisted of visual displays of four images, matched for style and size, combined with a six-word auditory sentence consisting of “*Now point to the*” followed by two critical words (an adjective and a noun). For example, participants might hear “*Now point to the blue car!*” when seeing a blue car, a gray car, a blue bottle, and a gray bottle; perceiving either critical word would be insufficient to identify the appropriate image.

Target sentences were presented either in isolation, or in the presence of one of three distractor signals: normal speech (an approximately six-word utterance made of concatenated individual words; e.g., “*Airplanes. Fly. High. And. Quite. Fast.*”), jumbled speech (the same words in a jumbled order; e.g., “*High. Fast. Fly. Airplanes. And. Quite.*”), or reversed speech (the normal speech sentence played in reverse; “*.tsaF. etuiQ. dnA. hgiH. ylF. senalpriA*”). All distractors were created by concatenating individually-recorded words (produced as part of a random word list) so as to avoid any coarticulatory or prosodic cues across words that would be disrupted when jumbling their order. The same word recordings were used in each of the three distractor conditions so as to match for spectral energy (EM). They differed primarily in the extent to which they provided linguistic information that could interfere with processing the target sentence (their potential for linguistic IM).¹ Target sentences were likewise produced as concatenated words in order to enhance opportunities for IM; the use of the same initial four words in all target sentences (“*Now. Point. To. The.*”) served to designate the target sentence and provided an opportunity for any buildup of stream-segregation and/or IM to develop. However, it is worth noting that the lack of

sentence-level prosody in our sentences may have made them more akin to word lists than true sentences, and could affect the results (a topic we return to later).

Six different female native-English speaking talkers were recorded, with each occurring equally often as target and as distractor. Talkers were provided with a list of approximately 125 words in alphabetical order to record in a noise-reducing sound booth, using a Shure SM81 microphone (Niles, IL) at a 44.1 kHz sampling rate and 16 bits precision, and adjusted to be the same peak amplitude. Words were concatenated (without intervening silence) to form the target and distractor sentences. All words in a single sentence were from the same speaker; target and background streams were from different talkers, began at the same time point, and were mixed together at equal root-mean-square intensity levels (0 dB signal-to-noise ratio). The background stream always ended slightly after the target, to ensure the final word in the target stream was not unmasked. These were presented diotically over headphones at a comfortable listening level (the same level for all participants).

Participants faced a computer running Psyscope X.² On each trial, they saw the four images appear and, simultaneously, heard a female talker tell them to point to one image (e.g., “Now point to the blue car!”). They first heard a 4-trial training block (one trial in each condition, with different sentences than in the test trials) to familiarize them with the task, and then proceeded to the test block, containing a total of 120 trials, with 40 per condition (quiet, normal-speech distractor, jumbled-speech distractor, and reversed-speech distractor), distributed equally among all combinations of target and background talker. Which sentence occurred in which condition was counterbalanced across participants; the location of the correct image (upper left, upper right, lower left, lower right) was counterbalanced across trials.

Adult participants were instructed to select the correct choice on a response box using one of four buttons; each trial progressed automatically to the next when the button was pressed. Children were asked to point to the correct image; the experimenter then pressed a button to move to the next trial. Children were instructed to guess when uncertain. All 120 trials were presented in random order in a single block, but for children pauses were taken to provide reinforcement stickers at approximately 15-trial intervals. The session was videotaped, and two trained observers coded all responses, with a third as tie-breaker in cases of disagreement. Accuracy was measured as a percentage score, but statistical analyses used arcsin transforms.

3. Results

As can be seen in Fig. 1, preschool-age children generally showed much poorer performance than the older children, who showed slightly poorer performance than adults. All three ages showed poorer performance with maskers present than without, and better performance with the reversed speech masker than with the intelligible maskers. To examine these effects statistically, we first conducted a 3 age \times 4 masker overall mixed effects analysis of variance (ANOVA), and (not surprisingly), found main effects of age [$F_{(2,53)}=122.39$, $p<0.0001$, $\eta_p^2=0.822$] and masker type [$F_{(3,159)}=75.83$, $p<0.0001$, $\eta_p^2=0.59$], as well as a significant interaction [$F_{(6,159)}=5.97$, $p<0.0001$, $\eta_p^2=0.18$]. Adults were more accurate (89%–95%) than school-age children [83%–97%;

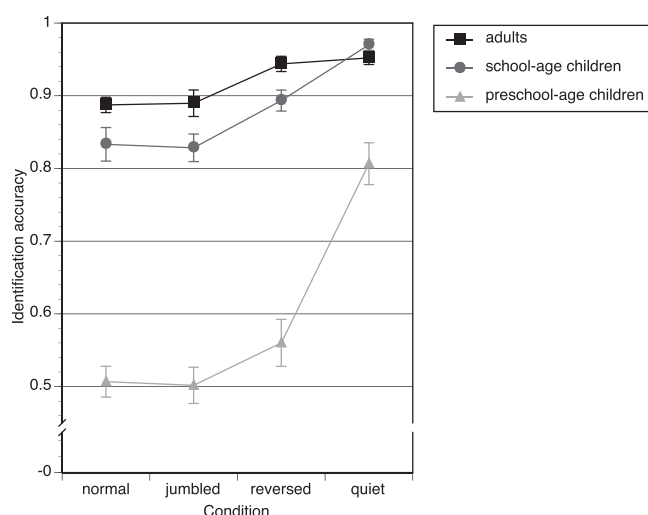


Fig. 1. Accuracy results (proportion of test words correctly identified) across conditions and listener groups.

$t_{(32)}=2.59$, $p=0.014$], who were in turn more accurate than preschool children [50%–81%; $t_{(38)}=11.84$, $p<0.0001$]; accuracy was much higher when there was no masker present (90%) than when there were any of the other three maskers (72%–79%). Given the interaction, we then explored in more depth the pattern of performance in each of the ages, starting with the oldest.

As expected, adult listeners showed a significant main effect of masker [$F_{(3,45)}=7.93$, $p<0.0001$, $\eta_p^2=0.346$]. Although this was driven in part by higher performance with no masker present (95.2%), the effect remained when only the three masking conditions were included in the analysis. Accuracy levels for normal (88.7%) and jumbled (89.0%) speech did not differ [$t_{(15)}=0.34$, $p=0.74$], but both differed from the reversed speech (94.4%) masker [normal speech: $t_{(15)}=4.22$, $p<0.0001$; jumbled speech: $t_{(15)}=2.69$, $p=0.017$]. Thus, adult listeners showed a significant release from masking when the masker was unintelligible, but did not show any release from masking when the masker consisted of interpretable words that did not make a legal sentence. This suggests that linguistic IM appears to be the result of lexical competition or source confusion, rather than of true sentence comprehension. It is worth noting, however, that our sentences did not contain natural sentence-level prosody; it is possible that this lack of prosody resulted in them not being treated as fully sentence-like; perhaps we would have found effects of sentence-level comprehension had there also been sentence-level prosody. Still, sentence-level meaning does not appear to add additional IM over that of individual words when divorced from prosody. Interestingly, adults showed very similar performance with a reversed speech masker as with no masker at all [$t_{(15)}=0.69$, $p=0.50$], providing no evidence for EM; when the masker did not contain any recognizable lexical items, it provided no apparent masking at all.

Like the adult listeners, school-age children showed a significant main effect of masker [$F_{(3,57)}=31.32$, $p<0.0001$, $\eta_p^2=0.622$]. Accuracy levels were nearly-identical for normal (83.3%) and jumbled (82.8%) speech maskers [$t_{(19)}=0.37$, $p=0.71$], but at higher levels for the reversed speech masker (89.3%). This is the same pattern as with adults. However, while the reversed-speech advantage was significant when compared to jumbled speech [$t_{(19)}=3.07$, $p=0.006$], it was only marginal when compared with normal speech [$t_{(19)}=2.07$, $p=0.053$]. The pattern of results suggests that school-age children, like the adult listeners, show a release from masking when the masking is unintelligible, and that this seems to be tied to lexical competition or source confusion, more than sentence comprehension.

The one major difference between school-age children and adults lies in a significant effect of masking in the reversed speech condition (89.3%) compared to the no-masker (97.0%) condition [$t_{(19)}=6.59$, $p<0.0001$]. Whereas 8-yr-olds performed similar to (or even slightly better than) adults in the control condition [$t_{(34)}=1.70$, $p=0.099$], they performed significantly more poorly in the reversed speech condition [$t_{(34)}=2.91$, $p=0.006$]. While they showed some release from masking for the reversed speech condition compared to other maskers, they did not show the complete release of masking shown by adults. Although school-age children appeared to perform more poorly than adults in the normal and jumbled speech conditions, neither difference reached significance [$t_{(34)}=1.33$, $p=0.19$ and $t_{(34)}=2.02$, $p=0.052$, respectively].

Preschool-age children showed much lower overall accuracy levels; comparing the two child groups, there are main effects of masker [$F_{(3,114)}=85.16$, $p<0.0001$, $\eta_p^2=0.69$] and age [$F_{(1,38)}=134.06$, $p<0.0001$, $\eta_p^2=0.779$], but no interaction [$F_{(3,114)}=1.57$, $p=0.201$, $\eta_p^2=0.04$]. This suggests that while the younger children performed more poorly overall, they showed the same general pattern as older children. Preschool-age children showed a main effect of masker [$F_{(1,19)}=57.25$, $p<0.0001$, $\eta_p^2=0.75$], with equivalent performance in the normal (50.7%) and jumbled (50.2%) conditions [$t_{(19)}=0.22$, $p=0.82$], but better performance (56.0%) in reversed speech [although this was significant only when compared to jumbled speech; reversed vs normal speech: $t_{(19)}=1.80$, $p=0.087$; vs jumbled speech: $t_{(19)}=2.11$, $p=0.048$]. Thus, even these preschoolers show a release from masking when the masker is unintelligible.

4. Discussion

Children must often attend to speech in settings that include multiple talkers speaking at once. These unattended/background talkers can interfere with the reception and comprehension of the speech of the target talker by causing either EM or IM or both. Children typically have more difficulty understanding speech in the presence of background talkers than do adults. At the outset, it was unclear whether this observation reflects a general reduction in performance under any type of multisource competition or whether it indicates that language has developed sufficiently to support

linguistically-based IM. Earlier work from Newman (2009) found no evidence for linguistically-based IM in infants. In the present study, we found that preschool-age children exhibited greater difficulty with the task in general, even with intelligibility in quiet. But in particular, their performance on the masked conditions was significantly lower than was found for older children or adults. This finding is consistent with prior studies demonstrating that children's performance is generally poorer than adults in noisy conditions (Fallon *et al.*, 2000; Hall *et al.*, 2012; Klatte *et al.*, 2010). However, although performance was generally poorer overall for the younger children, they nonetheless showed a pattern of masking as masker type varied that was very similar to that from adults. All three age groups showed better performance with an unintelligible time-reversed speech masker than an intelligible speech masker. Although this pattern of findings has been shown previously with adults (Hoen *et al.*, 2007; Summers and Molis, 2004), it has not previously been shown with young children (cf. Johnstone and Litovsky, 2006) or infants (cf. Newman, 2009).

The current findings with young children, similar to that with adults (e.g., Summers and Molis, 2004; Brouwer *et al.*, 2012), suggest that at least some of this IM may be the result of lexical competition between target and masker speech. Furthermore, the language development that presumably underlies this effect appears to be intact by 4 yrs of age (the mean age of our preschool group). For these procedures and subjects, generally poorer performance (than adults) on speech identification occurs concurrently with sensitivity to the linguistic content of competing speech maskers. In other words, while general listening ability remains immature at this age, the ability to show linguistic interference has matured. This interpretation presumes that the release from masking that occurred with reversed speech was the result of linguistically-based IM. Another possibility is that the pattern of results was driven by a release from EM caused by acoustic differences. While prior work has not examined EM with forward versus reversed speech in children, findings with adults suggest that the EM arising from both types of speech is generally equivalent. This work with adults supports our linguistic interpretation, yet further exploration of this topic with children is necessary to better tease apart the two possibilities.

Interestingly, none of the three groups showed any difference in performance levels when the background speech consisted of syntactically-correct sentences compared to jumbled words. This suggests that the effect of linguistic IM, as produced in this study, is the result of competition from activated lexical representations or from source confusion, and not the result of phrase or sentence processing. This is not to say that linguistic IM could not also be generated by sentence-level effects in another paradigm; our syntactically-correct sentences were produced by concatenating isolated words, and thus did not have the cohesive prosodic and coarticulatory patterns that would be found in typical sentences. It is possible that items with sentence-level prosody might induce processing at an additional syntactic level, one not seen here. But for the items presented here, IM appears not to be based on sentence-level meaning.

The present results suggest that if there is an age-related change in linguistic IM, it likely is occurring prior to 4 yrs of age. There are several possible explanations for why preschoolers, but not infants (as suggested by Newman, 2009), might show such an effect. Chief among them is the fact that linguistic IM requires the presence of lexical representations that can compete with the target word. Infants are likely to have fewer such lexical representations, and thus are less likely to demonstrate interference. Thus, the differences in IM may be the result of the number of stored lexical representations, and not the way that information is processed. One way to examine this hypothesis (that the processing that leads to IM is not age-dependent) would be to present young children with distractors consisting of words with lower vs higher typical age-of-acquisition (AOA); we might expect the same children to show linguistic IM for words with lower AOA but less masking for words with a higher AOA.

The current results conflict with the findings from Johnstone and Litovsky (2006) who reported that children showed the same amount of masking from reversed speech as from forwards speech, rather than the reduced masking reported here. They suggested that the "novelty of a reversed speech signal affected how children allocated their attention" (p. 2187); in effect, children were surprised by the unusual signal and attended to it rather than to the target signal. It should be pointed out, though, that the procedures used by Johnstone and Litovsky (2006) probably did not create much IM in the first place. This is because the target and masker talkers were well-segregated (a sex difference was used to identify the target distinct from the masker talkers), the target and masking speech were very different (spondees vs IEEE sentences) and so were not easily confused, and the masker began and ended before/after the

target enhancing segregation by temporal disparities. Consistent with this conclusion, they reported that a modulated noise masker caused greater masking than the intelligible speech masker. In contrast, here the target and maskers were same-sex talkers and the content was very similar likely leading to relatively high IM. However, there are a number of other differences between the two studies and thus the reason for the different outcomes is not certain.

The current results also do not match our initial predictions. We had expected that while school-age children would show a release from masking for un-interpretable speech, as do adults, that preschool children would not. Instead, we found that even 4-yr-old children demonstrated greater masking from an intelligible linguistic message. It is unclear if this could be assessed in still younger children without implementing significant methodological changes that could make comparisons across ages difficult. Still, the fact that this difference between time-reversed and forward masker speech is apparent as early as 4 yrs is a novel, and perhaps surprising, result.

Acknowledgments

This work was supported by NSF BCS1152109 to the University of Maryland, NIDCD R01 DC04545 to Boston University, and by an NRSA T32 and an NSF IGERT to UMD supporting G.M. Thank you to the Eleanor Roosevelt High School internship program, and to members of the Language Development Lab (especially K. Wilson, M. O'Fallon, J. Vesnovsky, and E. Slonecker) for assistance with recording, scheduling, and testing participants and data management.

References and links

¹Reversed speech also differs in temporal envelope structure, but research suggests this leads to greater (not lesser) masking (Rhebergen *et al.*, 2005), so any advantage for a reversed signal over an intelligible one should not be the result of these temporal changes.

²<http://psy.ck.sissa.it/>.

- Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (2012). "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," *J. Acoust. Soc. Am.* **131**(2), 1449–1464.
- Fallon, M., Trehub, S. E., and Schneider, B. A. (2000). "Children's perception of speech in multitalker babble," *J. Acoust. Soc. Am.* **108**(6), 3023–3029.
- Francart, T., van Wieringen, A., and Wouters, J. (2011). "Comparison of fluctuating maskers for speech recognition tests," *Int. J. Audiol.* **50**, 2–13.
- Hall, J. W., Buss, E., and Grose, J. H. (2012). "Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation," *Ear Hear.* **33**(3), 340–348.
- Hétu, R., Truchon-Gagnon, C., and Bilodeau, S. A. (1990). "Problems of noise in school settings: A review of literature and the results of an exploratory study," *J. Speech Lang. Pathol. Audiol.* **14**(3), 31–39. Available at http://209.217.105.25/english/resources/database/files/1990_JSLPA_Vol_14/No_03_1-67/Hetu_Truchon-Gagnon_Bilodeau_JSLPA_1990.pdf.
- Hoën, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., and Collet, L. (2007). "Phonetic and lexical interferences in informational masking during speech-in-speech comprehension," *Speech Commun.* **49**(12), 905–916.
- Hygge, S., Evans, G. W., and Bullinger, M. (2002). "A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren," *Psychol. Sci.* **13**(5), 469–474.
- Johnstone, P. M., and Litovsky, R. Y. (2006). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," *J. Acoust. Soc. Am.* **120**(4), 2177–2189.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science, New York).
- Klatte, M., Lachmann, T., and Meis, M. (2010). "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting," *Noise Health* **12**(49), 270–282.
- Newman, R. S. (2009). "Infants' listening in multitalker environments: Effect of the number of background talkers," *Attn., Percept., Psychophys.* **71**(4), 822–836.
- Oh, E. L., Wightman, F. L., and Lutfi, R. A. (2001). "Children's detection of pure-tone signals with random multitone maskers," *J. Acoust. Soc. Am.* **109**(6), 2888–2895.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2005). "Release from informational masking by time reversal of native and non-native interfering speech," *J. Acoust. Soc. Am.* **118**(3, pt. 1), 1274–1277.
- Shield, B. M., and Dockrell, J. E. (2003). "The effects of noise on children at school: A review," *Bldg. Acoust.* **10**(2), 97–116.
- Summers, V., and Molis, M. R. (2004). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," *J. Speech, Lang., Hear. Res.* **47**, 245–256.